Events Detection, Coreference and Sequencing: What's next?

Overview of TAC KBP 2017 Event Nugget Track

Teruko Mitamura
Zhengzhong Liu
Eduard Hovy
Carnegie Mellon University



TAC KBP Event Detection Tasks for English, Chinese and Spanish

- Goal: The task aims to identify the explicit mentioning of Events in text.
- 1.a. Event Nugget Detection TaskEvaluation Window: September 25 October 2
- 1.b. Event Nugget Detection and Coreference Task Evaluation Window: September 25 October 2
- Event Sequencing Task (English Only)
 Evaluation Window: October 3 -10



1.a. Event Nugget Detection Task for English, Chinese and Spanish

Participating systems will extract the following items:

- 1. Event Nugget Span Identification (character string)
- 2. Event Type and Subtypes (subset types of Rich ERE)
- 3. REALIS Value (one of: ACTUAL, GENERIC, OTHER)



1.b. Event Coreference Task for English, Chinese, and Spanish

- Input: Newswire and Discussion Forum documents (not annotated)
- Output: Event Nugget and Coreference Links
- Follow the notion of an Event Hopper (less strict coreference in ACE and light ERE)
- Corpus: Newswire and Discussion Forum



2015 TAC KBP EN tasks: 9 Event Types/ 38 Subtypes from Rich ERE Annotation Guidelines

- 1. Life Events (be-born, marry, divorce, injure, die)
- **2. Movement Events** (transport-person, transport-artifact)
- **3. Business Events** (start-org, merge-org, declare-bankruptcy, end-org)
- **4. Conflict Events** (attack, demonstrate)
- **5. Contact Events** (meet, correspondence, broadcast, contact)
- **6. Personnel Events** (start-position, end-position, nominate, elect)
- **7. Transaction Events** (transfer-ownership, transfer-money, transaction)
- **8. Justice Events** (arrest-jail, release-parole, trial-hearing, charge-indict, sue, convict, sentence, fine, execute, extradite, acquit, appeal, pardon)
- 9. Manufacture (artifact)



2016-2017 TAC KBP EN Tasks: 8 Event Types/18 Subtypes from Rich ERE Annotation Guidelines

- 1. Life Events (be-born, marry, divorce, injure, die)
- 2. Movement Events (transport-person, transport-artifact)
- 3. Business Events (start-org, merge-org, declare-bankruptcy, end-org)
- 4. Conflict Events (attack, demonstrate)
- 5. Contact Events (meet, correspondence, broadcast, contact)
- 6. Personnel Events (start-position, end-position, nominate, elect)
- 7. Transaction Events (transfer-ownership, transfer-money, transaction)
- 8. Justice Events (arrest-jail, release-parole, trial-hearing, charge-indict, sue, convict, sentence, fine, execute, extradite, acquit, appeal, pardon)
- 9. Manufacture (artifact)



REALIS Identification

- ACTUAL: the event actually happened
 - The troops are attacking the city. [Conflict.Attack,
 ACTUAL]
- GENERIC: the event is in general and not specific instance
 - Weapon sales to terrorists are a problem.
 [Transaction.Transfer-Ownership, GENERIC]
- **OTHER**: the event didn't occur, future events, desired events, conditional events, uncertain events, etc.
 - He plans to meet with lawmakers from both parties.
 [Contact.Meet, Other]



Evaluation for EN and Coreference

- Task 1.a: Event Nugget Detection (Span, Type, Realis, All)
 - English: 10 teams were submitted
 - Chinese: 3 teams were submitted
 - Spanish: 2 teams were submitted
- Task 1.b: Event Nugget and Coreference
 - English: 5 teams were submitted
 - Chinese: 2 teams were submitted
 - Spanish: 1 team was submitted



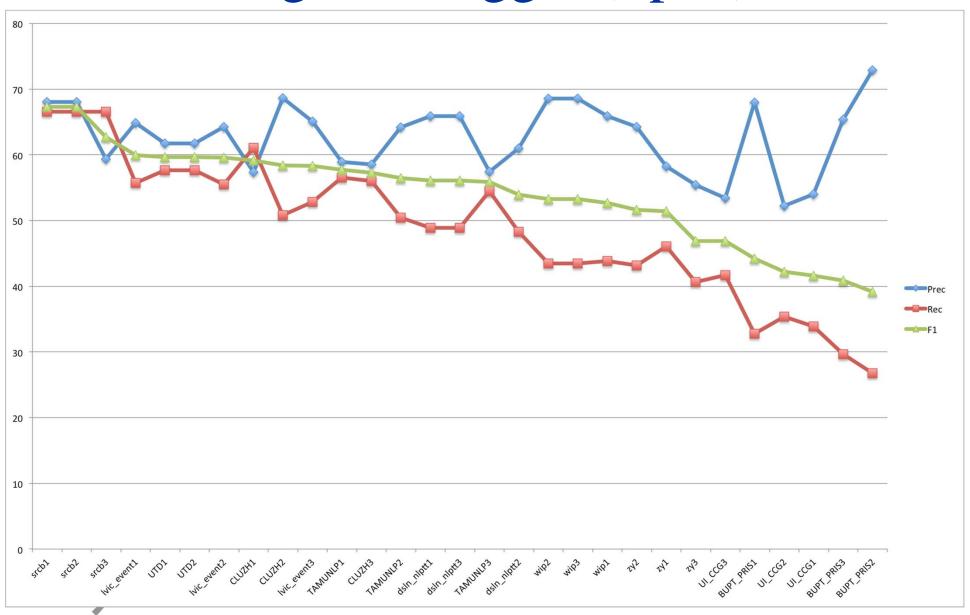
English
Nugget
Results
(Span)

Highest score from each team

	Prec.	Recall	F1
srcb1	68.04	66.53	67.27
lvic-event1	64.89	55.71	59.95
UTD1	61.74	57.66	59.63
CLUZH1	57.34	61.09	59.16
TAMUNLP1	58.95	56.53	57.72
dsln-nlptt1	65.89	48.87	56.12
wip2	68.58	43.5	53.24
zy2	64.29	43.14	51.64
UI-CCG3	53.44	41.72	46.86
BUPT-PRIS1	67.95	32.74	44.19



English Nugget (Span)



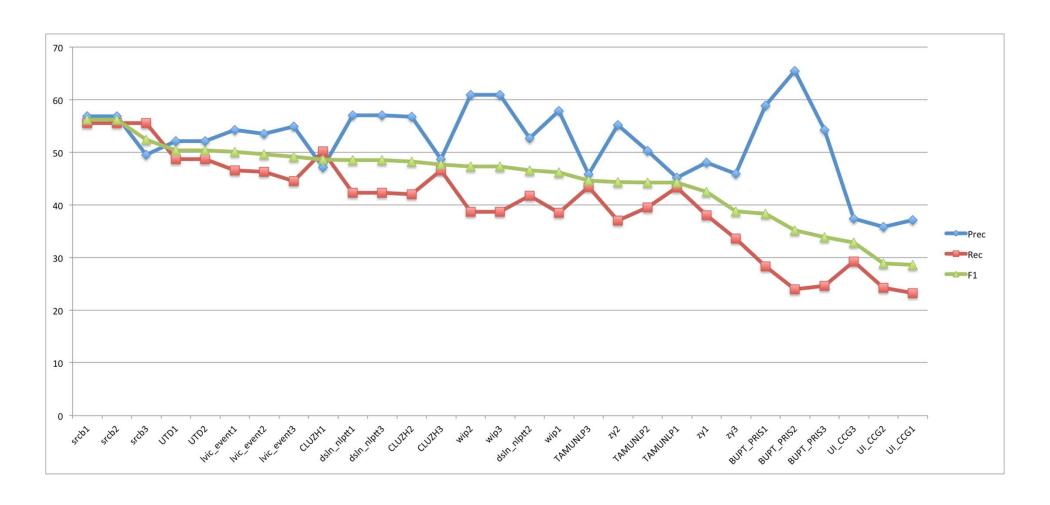
English
Nugget
Results
(Type)

Highest score from each team

	Prec.	Recall	F1
srcb1	56.83	55.57	56.19
UTD1	52.16	48.71	50.37
lvic-event1	54.27	46.59	50.14
CLUZH1	47.1	50.18	48.6
dsln-nlptt1	57.02	42.29	48.56
wip2	60.98	38.68	47.33
TAMUNLP3	45.88	43.48	44.65
zy2	55.22	37.06	44.35
BUPT-PRIS1	58.92	28.39	38.31
UI-CCG3	37.46	29.24	32.85



English Nugget (Type)





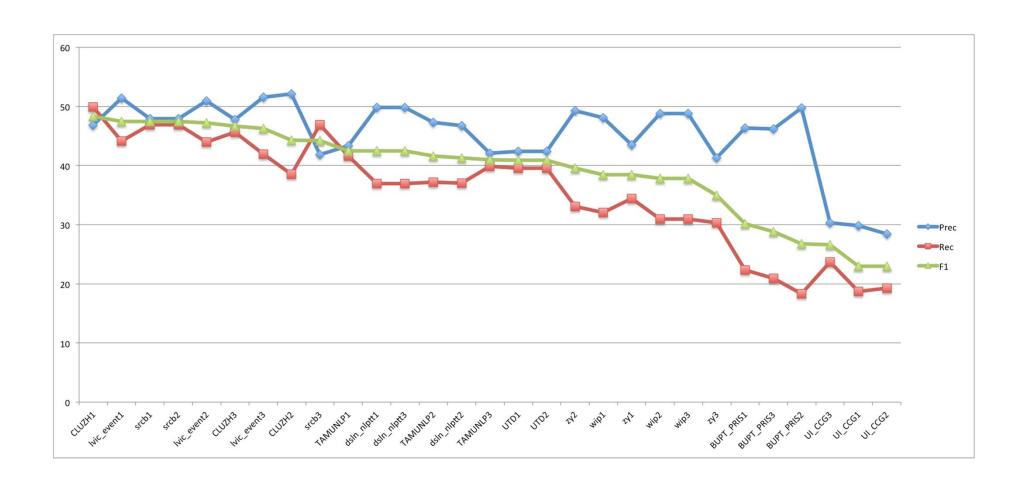
English
Nugget
Results
(Realis)

Highest score from each team

	Prec.	Recall	F1
CLUZH1	46.85	49.91	48.33
lvic-event1	51.39	44.12	47.48
srcb1	47.95	46.89	47.42
TAMUNLP1	43.38	41.6	42.47
dsln-nlptt1	49.86	36.98	42.47
UTD1	42.36	39.56	40.91
zy2	49.28	33.07	39.58
wip1	48.12	32.02	38.45
BUPT-PRIS1	46.36	22.34	30.15
UI-CCG3	30.3	23.65	26.57



English Nugget (Realis)





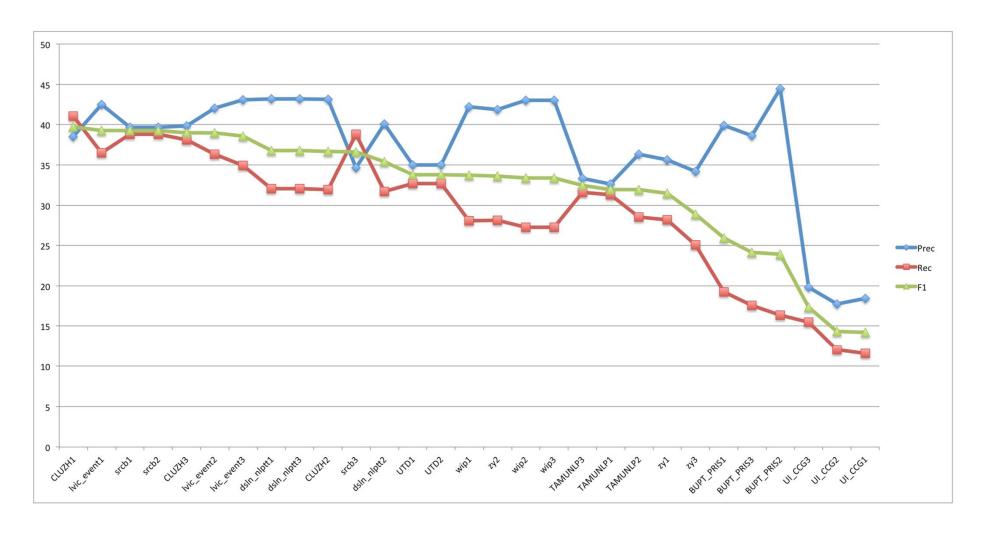
Task 1.a: English Nugget Results (All)

Highest score from each team

	Prec.	Recall	F1
CLUZH1	38.51	41.03	39.73
lvic-event1	42.52	36.5	39.28
srcb1	39.69	38.81	39.24
dsln-nlptt1	43.22	32.05	36.81
UTD1	35.01	32.7	33.81
wip1	42.21	28.08	33.73
zy2	41.87	28.1	33.63
TAMUNLP3	33.35	31.6	32.45
BUPT-PRIS1	39.92	19.24	25.96
UI-CCG3	19.8	15.46	17.36



Task 1.a: English Nugget (All)





Task 1.b: English Event Coreference

	B^3	CeafE	MUC	BLANC	Aver.
srcb2	43.84	39.86	30.63	26.97	35.33
UTD2	39.88	35.73	33.79	26.06	33.87
TAMUNLP2	34.34	33.63	22.9	17.94	27.2
BUPT-PRIS1	28.66	28.64	19.3	13.56	22.54
UI-CCG3	24.98	23.36	12.57	8.96	17.47



Observations on English Nugget and Coreference Tasks

- Most systems tend to have higher precision than recall.
- The best Event Nugget detection F1 score was 39.73, compared to 35.24 in 2016 and 44.24 in 2015.
- The best Event Type detection F1 score was 56.19, compared to 46.99 in 2016 and 58.41 in 2015.
- The best Event Coreference F1 score: 35.33, compared to 30.08 in 2016 and 39.12 in 2015.
- Part of the reasons may be caused by the reduction of Event Types/Subtypes to 18 from 38 in 2016 and many difficult and ambiguous event types remained: Transaction, Contact, etc.



Difficult English Event Types

- Contact-Broadcast, Contact-Contact, Transaction-TransferMoney, Transaction-TransferOwnership
- Transaction-TransferOwnership and Transaction-Transaction are easily misclassified.
- Movement-TrasnportArtifact was easily misclassified with Movement-TransportPerson.
- Contact-Broadcast was easily misclassified with Contact-Contact.



Chinese Nugget Results

Highest score from each team

		Prec.	Recall	F1
Span	CLUZH1	67.76	45.92	54.74
	UTD1	52.69	53.02	52.85
	srcb2	47.48	46.76	47.12
Type	CLUZH1	62.69	42.48	50.64
	UTD1	46.61	46.91	46.76
	srcb2	42.47	41.82	42.14
Realis	CLUZH3	49.66	38.5	43.37
	UTD1	35.08	35.3	35.19
	srcb3	34.87	34.3	34.58
All	CLUZH3	45.76	35.48	39.97
	srcb3	31.77	31.25	31.51
	UTD1	31.07	31.27	31.17



Results: Chinese Event Coreference

	B^3	CeafE	MUC	BLANC	Aver.
UTD1	34.18	32.22	27.07	18.57	28.01
srcb2	31.58	31.49	20.01	13.52	24.15



Spanish Nugget Results

		Prec.	Recall	F1
Span	CLUZH2	60.93	42.64	50.17
	UI-CCG3	37.4	26.62	31.1
Type	CLUZH2	51.99	36.38	42.81
	UI-CCG3	27.96	19.9	23.25
Realis	CLUZH1	45.63	30.85	36.81
	UI-CCG3	21.17	15.07	17.6
All	CLUZH1	38.36	25.94	30.95
	UI-CCG3	15.26	10.86	12.69



Spanish Event Coreference

	B^3	CeafE	MUC	BLANC	Aver.
UI-CCG1	9.9	10.39	3.89	2.04	6.55

- Only 1 team participated in Spanish
- The scores in Event Nugget and Coreference tasks are lower than English and Chinese.



Corpus Analysis



Event Coreference and Realis

Event sequence dataset in TAC KBP 2017 (extended by CMU)

		Train	Test
	# documents	360	169
	# event nuggets	15276	6124
	# Actual	9747 (63.8%)	3978 (65.0%)
	# Generic	2123 (13.9%)	390 (6.4%)
	# Other	3406 (22.3%)	1756 (28.7%)
	# singletons	8521 (55.8%)	3394 (55.4%)
	# non-singletons	6755 (44.2%)	2730 (44.6%)
7	# event clusters	2398	970

Exclude singletons



Event Coreference and Realis

- Event sequence dataset in TAC KBP 2017 (extended by CMU)
 - 'A only', 'G only', 'O only', and 'A & O' occupy 98-99%
 - 'A & G', 'G & O', and 'A, G & O' can be seen as misannotation (noise)

Exclude	sing	letons

<u>Legend</u>

A: Actual

G: Generic

O: Other

Train	Test
2398	970
1499 (62.5%)	629 (64.8%)
277 (11.6%)	56 (5.8%)
371 (15.5%)	206 (21.2%)
23 (1.0%)	4 (0.4%)
204 (8.5%)	72 (7.4%)
19 (0.8%)	3 (0.3%)
5 (0.2%)	0 (0.0%)
	2398 1499 (62.5%) 277 (11.6%) 371 (15.5%) 23 (1.0%) 204 (8.5%) 19 (0.8%)



Realis and Event Coreference

- He said he might attend the meeting. In fact, he attended it. [O, A] → Coref
- He said he might attend the meeting. However, he didn't attend it. [O, O] → Non-coref
- He said he might not attend the meeting. However, he attended it. [O, A] → Non-coref
- He said he might not attend the meeting. In fact, he didn't attend it. [O, O] → Coref
- The dog died. He did not live without food. [A, O] → Coref



- The 3-class distinction is not informative enough
 - The class 'Other' is too coarse-grained to differentiate affirmatives and negatives

Legend

[A]: Actual

[G]: Generic

[O]: Other

Event Sequence Task for English



Event Sequence Task for English

- Goal: Extract Subsequence of events within Doc
 - Input: Event nugget annotated files
 - Outputs: (1) After links; (2) Parent-Child links
- Corpus: Newswire and Discussion Forum in English
- Training data (After links and Parent-Child links were add by CMU to 2015 EN training and test data)
- Test Data creation by CMU
 - After links and Parent-Child links were add to 2016 EN test data
 - Event Nugget/Coreference links were added to the same
 Types/Subtypes as 2015 data sets, altogether there are 9 Event
 Types/ 38 Subtypes
- Annotation tool: Modified Brat tool
- Annotation Guidelines, Scorer, submission validation scripts and submission format were created by CMU



Two Types of Event-Event Relation Linking: AFTER Link and Parent-Child Link

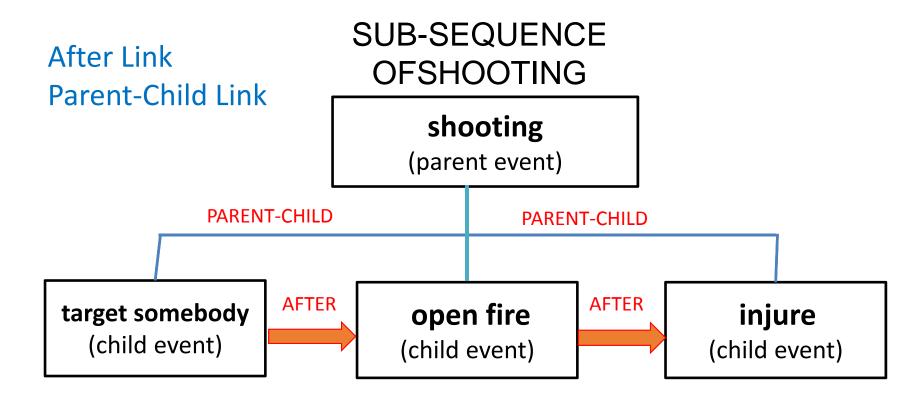
AFTER Link Relation:

- Represents a temporal sequence between child events in a subevent cluster
- Can be linked between child events with or without a parent event

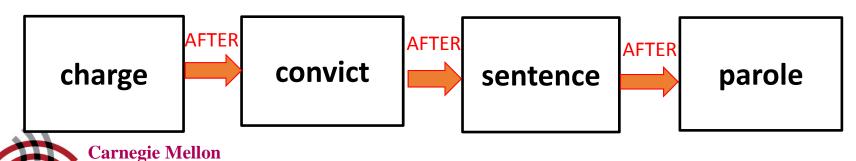
Parent-Child Link Relation:

Sub-event cluster detection





SUB-SEQUENCE OF JUDICIAL PROCESS



Language Technologies Institute

Event Sequence Task Results

- Only two teams submitted out of 16 teams registered
- After Link Detection (Top score)

	P	R	F1
KYOTOU	7.52	15.00	10.02

Parent-Child Link Detection (Top score)

	P	R	F1
KYOTOU	15.84	8.49	11.06



DEFT Pilot Study: Event Sequence Linking tasks for English

- Evaluation windows:
 - First: March 2-9, 2017
 - Second (informal): April, 2017
- Tasks: Extract sequence of events within doc
 - Input: Event nugget & coref annotated files
 - Outputs: (1) After links; (2) Parent-Child links
- CMU created:
 - Training data
 - Evaluation data
 - Annotation guidelines
 - Scorer
 - submission validation scripts and
 - submission format

Dataset

- Newswire and Discussion Forum in English from TAC KBP 2015
- Training set (N=157) LDC2015E73
 - 78 Discussion forum documents
 - 79 Newswire articles
- Evaluation set (N=202) LDC2015R26
 - 104 Discussion forum documents
 - 98 Newswire articles

DEFT English Event Sequence Pilot Study Results

The Evaluation was done in March-April, 2017.

	Subevent		After			
System	F1	Prec.	Recall	F1	Prec.	Recall
CMU1	-	-	-	18.5087	15.545	22.8688
CMU2	7.9083	9.4103	6.8198	17.5705	18.2877	16.9075
OSU1	10.979	12.5749	9.7425	14.4474	14.6508	14.2495
CMU3	7.3995	16.1215	4.8017	2.9427	8.2324	1.7916
UC1	-	-	-	3.7678	36.4532	1.9865



Issue 1: Granularity of Events

Example:

"Football: One dead after Croat and Muslim fans clash (E1)

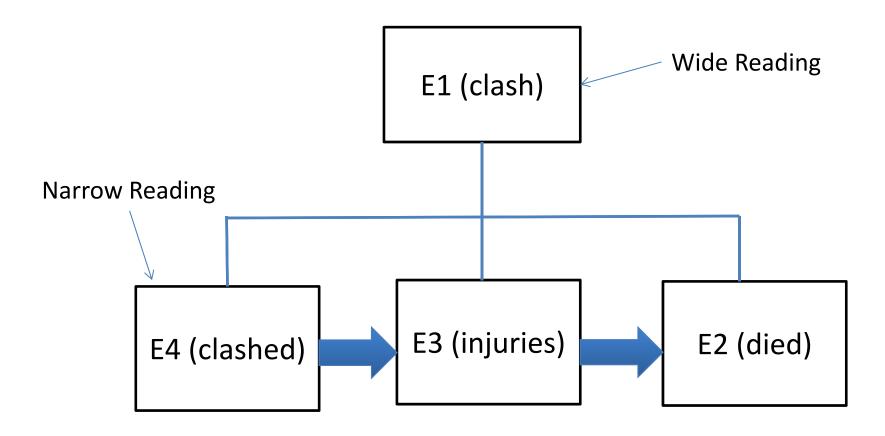
SARAJEVO, Oct. 4, 2009 (AFP)

One person died (E2) from injuries (E3) after Croat and Muslim fans clashed (E4) in the southern town Siroki Brijeg ahead of its Bosnian Premier League match (E5) against Sarajevo, police said (E6). [...]" (AFP ENG 20091004.0162)

- Granularity of events are sometimes determined by wide/narrow reading of events.
 - E1 (clash) is widely read to indicate the whole "clash" event.
 - E4 (clashed) is narrowly read to indicate the clash that occurred as part of E1 (clash).



Issue 1: Granularity of Events





Issue 2: Multiple Perspectives

- Events are sometimes reported from multiple perspectives. (e.g., testimonies in court)
- How do we sequence events which are reported from multiple perspectives?
 - Sequencing according to reporting agents
 - Akins' view
 - Kid Rock's view
 - Any other ways?



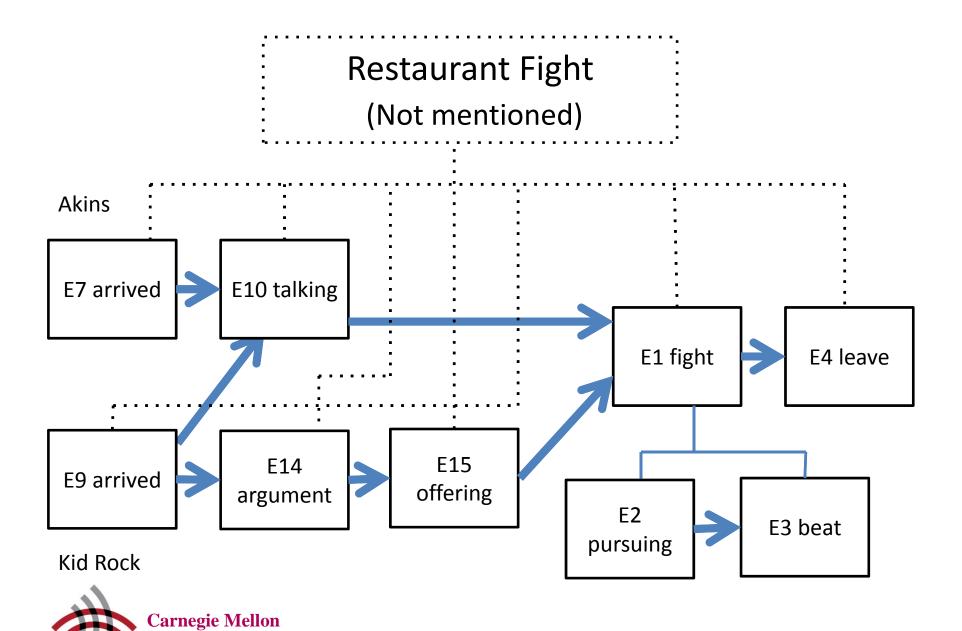
Example: Multiple Perspectives

The entertainer and his party behaved "like a pack of wild animals," starting a **fight (E1)** inside the restaurant and **pursuing (E2)** Akins into the parking lot to **beat (E3)** him up before **leaving (E4)** in their tour bus, Akins' lawyer Eric Hertz **said (E5)** in his opening **statement (E6)** in a DeKalb Country court. [...]

Akins arrived (E7) at the restaurant alone shortly after 5 a.m. local time on Oct. 21, 2007. Kid Rock, who had given a concert (E8) in Atlanta earlier, arrived (E9) in his tour bus around the same time. Akins and two women in Kid Rock's party, one of whom he had known for years, began talking (E10). Kid Rock was either jealous that Akins was getting the attention or was insulted by what Akins was saying (E11) to the women, but either way, a physical attack (E12) was unjustified, Hertz said (E13). Horton countered that Akins got into an argument (E14) with the women and with Kid Rock, who tried to calm things down by offering (E15) to buy Akins' breakfast.

(APW_ENG_20100914.0967)





Language Technologies Institute

What is next?

- Event Nugget Detection Task for English, Chinese, Spanish (Multilingual, Multi-Media, Cross-Doc)?
- 2. Full Event Coreference Task for English, Chinese, Spanish (Multilingual, Multi-Media, Cross-Doc)?
- 3. Even Sequence Linking tasks?
- 4. Open Domain EN tasks?



Questions?

